

# 记忆的结构

甄景贤 (King-Yin Yan)

General.Intelligence@Gmail.com

Abstract.

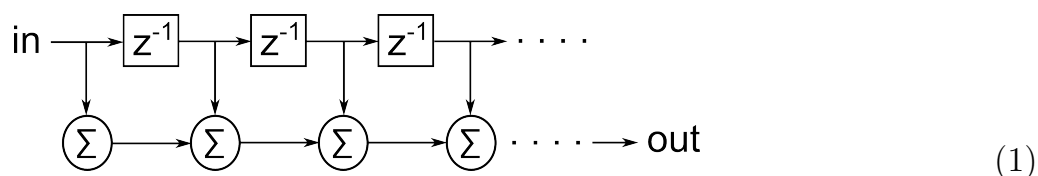
## 1 Episodic memory

设计了这个 minimalist architecture 之后，发现比起人脑有个严重缺陷：没有 **事件记忆** (episodic memory)。换句话说，只能留意当下发生的事件，但不能记住一段**故事**。

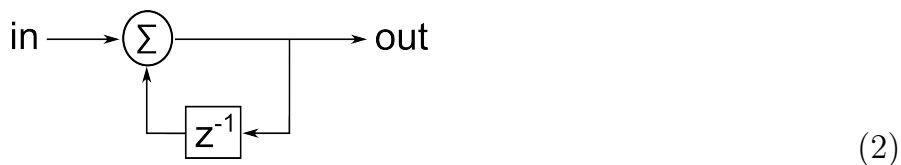
这牵涉到「什么是记忆？」的问题。在 minimal architecture 里， $F$  代表 “static knowledge”，亦即（相对地）**永恒不变**的知识 / 规律，而  $x$  代表当下的**状态** / 短期记忆，亦即 “dynamic knowledge”。Episodic memory 介乎「长期」与「短期」记忆之间。

**深度学习**在近年很厉害，但 deep NN 的缺点是没有**记忆**。有限自动机 (finite state machines) 不是**全能**的计算器，它和 Turing machine 的分别就是缺少了那条「记忆磁带」。所以，深度神经网络 + 记忆 就可以变成 universal 的计算器。如何设计「可微分」的记忆是一重要课题，因为可微分的记忆可以用 gradient descent 学习。

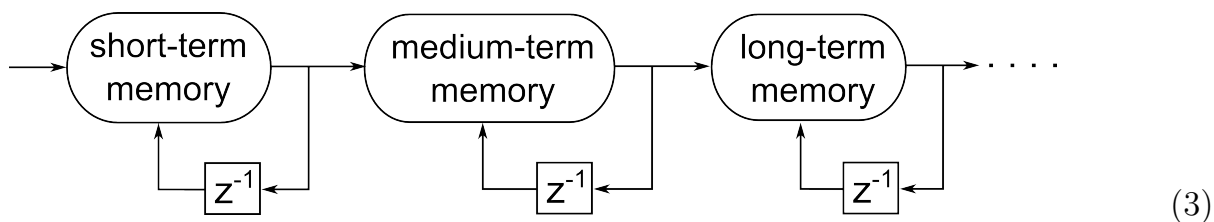
在 **信息处理** (signal processing) 理论中， $z$ -transform 以  $z^{-1}$  可以代表 **时间延迟** 1 步 (1-step time delay unit)。用一连串的  $z^{-1}$  可以组成长度为  $k$  单位的记忆：



以上的无穷串列可以用这个 recursive 结构代替：



多层的 hierarchical 记忆或许可以这样实现：



$Z$ -transform 的**连续时间**版本就是 Laplace transform。

上面的记忆结构来自 Haykin 的经典著作《Signals and Systems》[2]。

## 2 Mental state $x$ = working memory = 命题的集合

神经的状态空间由一些 **thoughts**（思维）组成，一个 thought 对应於逻辑中的一条命题，例如：

$$x = \text{我正在上课} \wedge \text{我很肚饿} \wedge \dots \quad (4)$$

这两个 thoughts 是独立的。也可以有另一个状态：

$$x_2 = \text{我正在搭地铁} \wedge \text{我很肚饿} \wedge \dots \quad (5)$$

Thoughts 独立的好处是表述的 economy（状态  $x$  分拆成若干独立的 thoughts）。 $x$  是  $M$  个 thoughts 的集合， $M$  是 working memory 的大小。认知科学里有个说法：

$$\text{the size of human working memory} \approx 7 \pm 2 \text{ items} \quad (6)$$

但这些 items 可以有 **chunking**，例如去超级市场买东西，「意大利粉、茄汁、芝士、香肠」这 4 件东西可以聚合成一个 chunk，这样可以记住的 items 数目多很多。

## 3 记忆 = 状态的历史

现在回看状态方程：

连续时间

 $\dot{x} = f(x) \quad (7)$

$\dot{x}$  是状态  $x$  的改变方向，换句话说，这是描述状态变化的方程。但状态的变化  $f(x)$  不取决於状态的历史（ $x = x|_t$  仅代表状态在时间  $t$  的值），所以这个系统没有记忆。如果想要记忆的话，一个简单的做法是令：

状态  $x$  的历史  


$$\dot{x} = f(x|_t, \underbrace{x(t \in \mathbb{T})}_{\text{状态 } x \text{ 的历史}}) \quad (8)$$

$f$  就是我们的神经网络。 $f$  的输入是：

- $x|_t$  = 当下的状态 = 一些命题的集合；命题没有 time stamp
- $x(t \in \mathbb{T})$  = 状态的历史；要考虑每个状态的 time stamp

## References

1. Simon Haykin. *Cognitive dynamic systems*. Cambridge Univ Press, 2012.
2. Simon Haykin and Barry van Veen. *Signals and systems*. John Wiley & Sons, 2003.